Use of untargeted metabolomics for assessing soil quality and microbial function

Withers, Emma; Hill, Paul W.; Chadwick, David R.; Jones, Davey L.

Soil Biology and Biochemistry

DOI:
10.1016/j.soilbio.2020.107758

Published: 01/04/2020

Peer reviewed version

Use of untargeted metabolomics for assessing soil quality and microbial function

Emma Withers[a,*], Paul W. Hill[a], David R. Chadwick[a,b], Davey L. Jones[a,c]

[a] *School of Natural Sciences, Environment Centre Wales, Bangor University, Gwynedd, LL57 2UW, UK*

[b] *Interdisciplinary Research Centre for Agriculture Green Development in Yangtze River Basin, Southwest University, Chongqing, China*

[c] *SoilsWest, UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA 6009, Australia*

*Corresponding author.

E-mail address: e.withers@bangor.ac.uk (E. Withers)

ABSTRACT

Soils support a wide range of ecosystem services that underpin Earth system functioning. It is therefore essential that we have robust approaches to evaluate how anthropogenic perturbation affects soil quality and the delivery of these services. Metabolomics, the large-scale study of low molecular weight organic compounds in soil, offers one potential approach to characterise soils and evaluate the metabolic status of the soil biological community. The aims of the present study were to 1) characterise the soil metabolome across a contrasting range of soil types, 2) understand the relationships between common chemical and physical soil quality indicators and its metabolome, and 3) evaluate the discriminatory power of soil metabolomics and its potential use as a soil quality indicator. Nine different topsoils with 5 replications were collected along an altitudinal primary productivity gradient encompassing a wide range of soil types and land uses. Metabolites were extracted from soil using 3:3:2 (v/v/v) acetonitrile:isopropanol:water and individual compounds identified using a gas chromatography-mass spectrometry (GC-MS) platform. Overall, 405 individual compounds were detected, of which 146 were positively identified, including sugars, amino acids, organic acids, nucleobases, sugar alcohols, lipids and a range of secondary metabolites. The concentration and profile of metabolites was found to vary greatly between the soil types. Further, the soils' metabolomic fingerprints correlated to a number of environmental factors, including pH, land-use, moisture and salinity. We also tentatively attributed soil-specific metabolites to potential functional pathways, although complementary proteomic, genomic and transcriptomic approaches would be needed to provide definitive supporting evidence. In conclusion, soil metabolomics offers the potential to reveal the complex molecular networks and metabolic pathways operating in the soil microbial community and a means of evaluating soil function. Further work is now required to benchmark soil metabolomes under a wide range of management regimes so that they can be used for the quantitative assessment of soil quality.

Keywords: Biomarker; Chemical fingerprinting method; Metabolic profiling; Microbial function; Soil health indicator.

1. Introduction

Soils are central to a wide range of ecosystem services that are essential to earth system functioning (Bünemann et al., 2018). It is therefore essential that we monitor the health of our soils so that the delivery of ecosystem services can be maintained (e.g. nutrient cycling, water purification, food provisioning, climate regulation). While a range of soil quality indicators have been proposed, these are mainly focused on the measurement of standard chemical attributes of the soil (e.g. pH, available P and K, organic matter content) and the physical characteristics of the soil (e.g. texture, structure, aggregate stability, bulk density; Schloter et al., 2018). However, soil fertility and productivity are not solely a function of the soil's physical and chemical characteristics. Soil organisms are key mediators of many processes linked to plant health and soil productivity. Despite many attempts, the development of robust soil biological quality indicators that can be widely adopted has remained elusive (Schloter et al., 2018). Examples of traditional indicators include measurements of biological activity (e.g. basal and substrate-induced respiration, enzyme activity) and the size and composition of the microbial community (e.g. $CHCl_3$ fumigation-extraction, fatty acid biomarkers) (Bending et al., 2004). The advent of 'omic'-based technologies aimed at the universal detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics), however, offers new ways to evaluate soil biological functioning. While the use of metagenomics and metabarcoding is becoming mainstream (Fierer et al., 2003; George et al., 2019), much less attention has been paid to the metabolomic profiling of soil microbial communities.

Untargeted metabolomics allows a global analysis of the low molecular weight (< 1000 Da) metabolites present within a sample (Vinayavekhin and Saghatelian, 2010). Through recent advances in spectroscopy, it is now feasible to identify and quantify the relative abundance of thousands of metabolites present in biological samples (Patti et al., 2012). A metabolomic approach is similar in cost to genomics and proteomics (Wilson et al., 2005), allows for rapid sample processing (Jones et al., 2013) and is not restricted by unknown degrees of epigenetic regulation and post-translational modifications, respectively (Patti et al., 2012). Additionally, the technique has the capacity to identify biochemical intermediates in interacting metabolic pathways, potentially improving our overall understanding of biological processes operating in soil and improving our ability to predict outcomes (Tang, 2011).

Applications of metabolomics within the environmental sciences extend from organism phenotype characterisation (Bingol et al., 2016; Patti et al., 2012); assessment of responses of plant and soil organismal assemblages to biotic and abiotic factors (Bundy et al., 2003, 2009; Jones et al., 2013, 2014; Trauger et al., 2008); characterisation of differential microbial community structures (Abram, 2015; Graham et al., 2018); and biomarker discovery (Bundy et al., 2009). Combined with complementary '-omics' techniques (genomics, proteomics, transcriptomics), metabolic profiling can provide a better overall understanding of molecular mechanisms associated with environmental cues (Trauger et al., 2008). Applied to the soil microbiome, metabolomics may provide a means of characterising the differential activity of microbial communities (Abram, 2015), reflecting microbial genome-environment interactions (Tang, 2011) and thus a novel way to assess soil health. This can be used to improve our understanding of cellular pathways and community responses to abiotic and biotic stress events as well as providing insights on fundamental soil biochemical functioning (Abram, 2015; Patti et al., 2012; Swenson et al., 2015).

The contribution of complex biological factors, such as soil microbial diversity, and the extent to which this provides functional redundancy in terms of ecosystem service provision, remains relatively unknown (Jurburg and Salles, 2015). Further, little is understood about the soil microbial metabolome, and the degree to which metabolomic fingerprints of soil classes may differ. Untargeted metabolomics analysis may therefore provide a means of assigning phenotype to specific metabolite expression (Guijas et al., 2018); identifying soil-specific microbial nutrient and cellular pathways; and attributing corresponding biological mechanisms and function (Patti et al., 2012; Zhao et al., 2019). Therefore, metabolomics could prove very useful in the assessment of how land use change, climate perturbation and land management regime affects soil health. In this context, the aims of this study were to apply untargeted metabolomics coupled with chemical characterisation to: 1) characterise the soil metabolome across a contrasting range of soil types and land uses, 2) understand the relationships between major chemical and physical characteristics of the soil and its metabolome, 3) evaluate whether metabolomics can provide a suitable indicator of soil quality.

2. Materials and methods

2.1. Soil sampling

Nine sites with different combinations of soil type and/or vegetation cover were sampled at the start of the growing season (March 2018) across a 350 m altitudinal gradient (catena sequence) at the Henfaes Experimental Station, Abergwyngregyn, UK (53°14'N, 04°01'W; Fig. S1). The sequence of nine soil types along the altitudinal gradient (from 0 to 350 m) were: Saline Alluvial Gley Soil 1, Saline Alluvial Gley Soil 2, Gleyic Sandy Brown Soil, Typical Orthic Brown Soil, Stagno-Orthic Gley Soil, Typical Podzolic Brown Soil 1, Typical Podzolic Brown Soil 2, Typical Humic Ranker Soil and Non-Calcaric Lithosol. The soils were classified on site according to the UK system of Avery (1990). The major properties of the

sites and soils are shown in Table 1 and in Figure S1, while a general description of the catena sequence is provided in Farrell et al. (2014). The altitudinal gradient also constitutes a primary productivity gradient with more intensive agricultural production at low altitudes. The mean annual temperature at the bottom and top sites was 10.2 and 7.3 °C respectively, while the gradient in annual rainfall was 1065 to 1690 mm, respectively. All sites had a different vegetation cover (all dominated by grasses) and were grazed by sheep (*Ovis aries* L.). Land boundaries within which each of the 9 discrete soil types was independently present were identified. Within each boundary, five randomly located independent 5 cm diameter soil cores (10 cm depth) were removed using a stainless-steel corer and placed in plastic bags. A fixed sampling depth was chosen to reflect national soil monitoring programmes (Bellamy et al., 2005; Emmett et al., 2008). Immediately after collection, the central 1 cm^3 was isolated from each core using a sterile spatula, the roots removed and the samples stored in sterile tinfoil cups at -80°C to await metabolome analysis. The remaining soil was retained, placed in plastic bags and stored at 4 °C for further analysis of the soil properties.

2.2. Untargeted metabolomics

The 45 collected soil samples, and 5 blank samples containing no soil, were lyophilized on an Edwards Super Modulyo freeze-drier (SciQuip Ltd., Shropshire, UK) for 7 d. Subsequently, these were ground in a Retsch MM200 stainless steel ball mill (Retsch GmbH, Haan, Germany) at a frequency of 20 Hz to aid recovery of metabolites from the microbial biomass (Fiehn et al., 2002; Wang et al., 2015). The samples were then stored in individual sterile glass vials at -80°C to minimize changes in metabolites (Wellerdiek et al., 2009). The soils were extracted using 3:3:2 (v/v/v) acetonitrile-isopropanol-water, vortexed for 15 seconds, shaken at 4ºC for 5 minutes, centrifuged at 1400 rpm for 2 minutes, and dried using a CentriVap Benchtop Centrifugal Concentrator (Labconco Corp., Kansas City, MO) (Barupal

et al., 2019; Fu et al., 2019). Non-targeted primary metabolism analysis was performed using a Gerstel Automated Linear Exchange-Cold Injection System (ALEX-CIS) with Agilent gas chromatograph (GC) and Leco Pegasus IV Time Of Flight (TOF) MS at the UC Davis West Coast Metabolomics Facility using the method of Fiehn (2016). Briefly, 0.5 µl of each sample was injected onto a Rtx-5Sil MS capillary column (30 m length × 0.25 m i.d with 10 m integrated guard column; 0.25 µm 95% dimethylsiloxane/5% diphenylpolysiloxane coating; Restek Corp., Bellefonte, PA). Using a He mobile phase, the GC thermal programme was 50 °C for 1 min, ramped to 330 °C at 20 °C min^{-1} and finally held at 330 °C for 5 min. Upon elution, samples were injected into a Pegasus IV GC-time of flight mass spectrometer (Leco Corp., St Joseph, MI), using mass resolution of 17 spectra s^{-1}, from 80-500 Da, at -70 eV ionization energy and 1800 V detector voltage with a 230 °C transfer line and 250 °C ion source.

2.3. General soil properties

Soil pH and electrical conductivity (EC) were measured in 1:2.5 (w/v) soil-to-distilled water extracts using standard electrodes. Moisture content was measured gravimetrically by oven drying (105 °C, 16 h). Available ammonium and nitrate were determined colorimetrically in 1:5 (w/v) soil-to-0.5 M K_2SO_4 extracts using the salicylic acid procedure of Mulvaney (1996) and vanadate procedure of Miranda et al. (2001), respectively on a Synergy® microplate reader (BioTek Instruments Ltd., Winooski, VT). Total free amino acid concentration in the 0.5 M K_2SO_4 extracts was determined fluorometrically using the *o*-phthalaldehyde-β-mercaptoethanol method of Jones et al. (2002). Available P was determined colorimetrically in 1:5 (w/v) soil-to-0.5 M acetic acid extracts using the molybdate blue method of Murphy and Riley (1962). Exchangeable Ca, Na and K in the 0.5 M acetic acid extracts was determined using a Model 410 flame photometer (Sherwood Scientific Ltd, Cambridge, UK). Total C and

N were determined on a TruSpec® CN analyser (Leco Corp., St Joseph, MI). Dissolved organic C (DOC) and total dissolved N (TDN) in the 0.5 M K_2SO_4 extracts were determined using a Multi NC 2100S TOC TN analyzer (AnalytikJena, Jena, Germany).

To measure substrate-induced respiration, field moist, root-free soil (5 g) was placed in individual 50 cm^3 polypropylene tubes. Subsequently, 1 ml of a ^{14}C-labeled glucose solution (1 mM; 1.6 kBq ml^{-1}) was added to the soil surface. A vial containing 1 M NaOH (1 ml) was then suspended above the soil to capture any $^{14}CO_2$ evolved and the tubes hermetically sealed and incubated at 20 °C. The NaOH traps were replaced after 0.5, 1, 2 and 4 h. After removal, the NaOH was mixed with Optiphase HiSafe 3 scintillation cocktail (PerkinElmer Inc., Waltham, MA) and the ^{14}C quantified on a Wallac 1404 liquid scintillation counter with automated quench correction (Wallac EG&G, Milton Keynes, UK). The procedure described above was repeated using ^{14}C-labelled maize leaf (50 mg) in place of the ^{14}C glucose. In this case the NaOH traps were replaced after 3 d (Simfukwe et al., 2011). The turnover of glucose and leaf material were subsequently referred to as labile and more recalcitrant C, respectively.

2.4. Data and statistical analysis

The metabolomics data were pre-processed using ChromaTOF (v2.34; Leco Corp.). Briefly, subtraction of the baseline was applied just above the noise level and automatic mass spectral deconvolution and peak detection applied at a 5:1 signal-to-noise ratio throughout the chromatogram. A BinBase algorithm (rtx5) was applied, spectra were cut to 5% base peak abundance and matched to database entries. Unmatched peaks were entered as new database entries where the signal-to-noise ratio was >25 and purity <1.0.

The data was normalized by $\log_{10}$ transformation for all subsequent analysis. Principal Component Analysis (PCA) was applied as an unsupervised method of determining variance within and between soil classes. A pairwise score plot was generated to determine the most

appropriate combination of Principal Components (PC) to include in the 2D score plot. Biplots were generated to visualise the contribution of the loading of each metabolite towards observed variance in the data.

Agglomerative hierarchical clustering analysis was applied to metabolite concentration data and soil classes using two separate methods. Firstly, similarity was determined by Euclidean distance for analysis of the differences in metabolite concentrations, and clustering was performed using Ward's linkage. Secondly, similarity was determined by Pearson's correlation for analysis of the shapes of metabolite expression profiles, and clustering was again performed using Ward's linkage. The dendrograms were combined with a heatmap, generated based on *z*-scores of metabolite concentrations.

A one-way ANOVA coupled with Fisher's LSD method was used to identify significant differences between metabolite concentrations in soil types using a $P < 0.05$ cut-off value to denote statistical significance. The same method was applied to identify significant differences between general soil properties observed in different soil types.

3. Results

3.1. Metabolic profile analysis

Using the methods described, 405 individual metabolites were detected across the nine distinct soil types sampled. 136 metabolites were observed in Saline Alluvial Gley Soil 1 in significantly higher concentrations than the blank control sample ($p < 0.05$), 181 in Saline Alluvial Gley Soil 2, 209 in Gleyic Sandy Brown Soil, 143 in Typical Orthic Brown Soil, 232 in Stagno-orthic gley soil, 253 in Typical Podzolic Brown Soil 1, 256 in Typical Podzolic Brown Soil 2, 253 in Typical Humic Ranker Soil, and 319 in Non-Calcaric Lithosol (Fig. S2). 146 of the 405 detected metabolites were positively identified (36% of the total), while 259 showed no match to spectra in the ChromaTOF database. Where PCA was applied to observe

variance within and between individual soil types, the combination of PC 1 and 2 offered best class separation compared with all other combinations of PCs (Fig. S3). PC 1 and 2 were therefore used to generate 2D PCA scores plots, which separated the nine different soil types into four distinct groupings (Fig. 1, Fig. S4):

Group 1. Non-Calcaric Lithosol (NC Lithosol).

Group 2. Saline Alluvial Gley Soil 1 (Saline 1) and Saline Alluvial Gley Soil 2 (Saline 2).

Group 3. Typical Orthic Brown Soil (TO Brown).

Group 4. Gleyic Sandy Brown Soil (Gleyic Sand), Stagno-Orthic Gley Soil (SO Gley), Typical Humic Ranker Soil (Humic Ranker), Typical Podzolic Brown Soil 1 (Podzolic 1) and Typical Podzolic Brown Soil 2 (Podzolic 2).

Within the fourth group, a significant difference in variance was observed between the Gleyic Sand and Podzolic 1, and between the Gleyic Sand and Humic Ranker soils. No significant differences in variance could be observed between any other soil types within this group. The majority of metabolites showed strong positive loadings in PC1. The TO Brown soil separated from all the other soil types by a lack of significant contribution from any specific metabolites (Fig. 1, Fig. S5). Both unassigned metabolite 250754 and pipecolinic acid (PIP) contributed significant loadings in the direction of Saline 1 and Saline 2, with the former doing so to a much greater extent (Fig. S5).

3.2. General soil properties

In contrast to the metabolite profiles, PCA of general soil properties (Fig. 2) generated just two distinct clusters. Together, Saline 1 and 2 showed significant difference in variance from all other soil types. The remaining seven soil types clustered together, however, a significant difference in variance could be observed between the Gleyic Sand and Podzolic 2,

and between the Humic Ranker and Podzolic 2. No further discrimination between soil classes could be made through this model.

General soil properties which segregated soil classes included the rate of recalcitrant C turnover, which generally decreased up the hillslope (Table 1). Soil pH also generally decreased with increasing altitude: Saline 1 and Saline 2 were alkaline (pH 7 to 9); Gleyic Sand, TO Brown, SO Gley and Podzolic 2 were circum-neutral (pH 5 to 7); and Podzolic 1, Humic Ranker and NC Lithosols were acidic (pH 3 to 5). Total C and N, moisture content and labile C turnover rate generally increased up the hillslope (Table 1). A similar trend was observed for NH_4^+ content in all soils other than the NC Lithosol, while no clear trend was seen for NO_3^-.

3.3. Metabolite concentration

Based on ANOVA of all detected metabolites (405 in total), significant differences ($p < 0.05$) were observed between mean concentrations of 344 metabolites in each of the nine distinct soil types (i.e. 86% of the total; Fig. S6). Where only the 146 known metabolites were considered, a significant difference ($p < 0.05$) was observed between mean concentrations of 140 metabolites (i.e. 96% of the total; Fig. S7).

Heatmaps (Fig. 3) detail the expression profiles of each soil sample by metabolite concentration *z*-score, based on the top 50 most significant known metabolites assigned by ANOVA (Fig. S7). Clustering metabolites and soil samples by Pearson correlation and Ward's linkage (Fig. 3A) distinguished the Saline 1, Saline 2 and NC Lithosol soils from one another. The same methods distinguished these three soils, within a cluster (Cluster A^1), from all other soil types. Podzolic 1 and Humic Ranker existed within a single cluster (Cluster A^2) largely defined by metabolites N-acetyl-D-hexosamine (Nah) to phosphate (excluding undecaprenyl phosphate N-acetylglucosamine; UDP-GlcNAc) being present at the highest concentrations,

although one Humic Ranker sample existed in Cluster A^1. Gleyic Sand, TO Brown, SO Gley and Podzolic 2 constituted Cluster A^3. Soils in Clusters A^2 and A^3 were more closely related to one another than to Cluster A^1 soils. When metabolites and soil samples were clustered by Euclidean distance and Ward's linkage, a different pattern was observed in comparison to method A, producing 6 distinct clusters (Fig. 3B). Briefly, Saline 1 and Saline 2 were clustered independently from one another, within a cluster (Cluster B^2). Occurring within Group 1, these soils were most closely related to Cluster B^1 soils (TO Brown). NC Lithosols clustered independently (Cluster B^3), whilst Podzolic 1 and Humic Ranker soils clustered together (Cluster B^4). Cluster B^4 was linked most closely with Group 2, containing Gleyic Sand (Cluster B^5), and SO Gley and Podzolic 1 (Cluster B^6).

4. Discussion

4.1. Do soil types possess unique metabolomic fingerprints?

Our results revealed a wide variation in metabolite concentration across the altitudinal gradient. The similarity between the metabolomic profiles for some soil type/vegetation combinations (Fig. 1), however, indicated that each site was not unique. This is similar to genomic-based measurements of soil microbial (e.g. bacteria, fungi, archaea) and mesofaunal communities which also showed that some of these soil types do not possess unique biological fingerprints (George et al., 2019). In this latter study, separation in communities was more related to vegetation cover, soil pH and organic matter content than soil type *per se*. Metabolomic profiling by GC-MS therefore does not appear to provide a means of uniquely defining soils, but does allow clustering of soils with similar biochemical properties. As metabolomic analysis provides a signature of functional metabolic processes (Bundy et al., 2009; Patti et al., 2012), our results support the view that considerable functional redundancy exists across soil classes. This is consistent with the view that (i) the primary metabolism of

soil organisms is likely to be similar, irrespective of soil type, and (ii) many soils share a common core microbiome, particularly when the vegetation cover is similar (Bergmann et al., 2011; Barberan et al. 2012; Jones et al., 2018). It is possible that separation on secondary, rather than primary, metabolites might facilitate greater sample separation, however, this requires an advancement in analytical capability. Although we quantified 405 individual metabolites, this probably represents a tiny fraction of the low molecular weight compounds actually present in our soils. For example, in animal- and plant-based metabolomic studies the number of compounds identified can be >1000 (Huan et al., 2016; Mahieu and Patti, 2017), suggesting the need to improve the extraction efficiency of solutes from soil and to pre-concentrate them prior to analysis. Continual advances in GC-MS analytical resolution and chemical reference libraries should also enhance the resolution of the technique and reduce the proportion of unknown compounds (Mahieu and Patti, 2017; Wishart, 2019).

The observed similarity between some soils may also partly reflect the fact that we only studied topsoils. Typically, soil classification systems use both diagnostic topsoil and subsoil characteristics (e.g. gleyed or podzolic B horizons). It is therefore recommended that the sampling of multiple soil horizons be undertaken to increase the potential to discriminate between soil types.

4.2. Does metabolomics provide greater discriminatory power than conventional soil quality indicators?

Traditional soil quality indicators (SQIs) allowed us to segregate the nine soils into just two distinct groups whereas metabolomics identified four distinct clusters (Figs. 1-2), indicating that metabolomics provides greater classification power. The clusters we identified were also similar to those determined from microbial substrate use profiles across 500 sites (Simfukwe et al., 2011). A cluster analysis of 1350 sites across Wales using traditional SQIs

also revealed significant crossover between soil types but identified 4 distinct soil groupings, based mainly on pH and organic matter status, namely: organic, organo-mineral, acid mineral, and neutral mineral soils (Seaton et al., 2019). Although we did not have any organo-mineral soils, these are also relatively consistent with our groupings: the > 15% TC coupled with a pH range 3-5 observed in 90% of organic soils matches the 29.14% TC and pH 4.27 measured in NC Lithosol; the <11 % TC coupled with a pH range 4.2 - 7.5 observed in 90% of neutral mineral soils matches the 3.62 % TC and pH 5.78 measured in TO Brown; and the < 11% TC coupled with a pH range 4.3 – 6.8 observed in 90% of acid mineral soils approximately matches the 2.64 – 11.57% TC and pH 4.37 – 5.68 measured in Gleyic Sand, Podzolic 1, Podzolic 2 Humic Ranker and SO Gley (Table 1, Fig. 1). Of note is that this previous study did not capture the saline grouping identified in our study.

Separation between Clusters B^4 and B^5 with Cluster B^6 through agglomerative clustering by Euclidean distance (Fig. 3) contrasts with the lack of significant differences in variance observed between the respective soil classes through PCA (Fig. 1). Based on PCA, the metabolomic profiles of the members of each cluster overlaps, however, the dendrogram indicates that metabolite concentrations of within-cluster soils are more similar to one another than to metabolite concentrations of soil classes in different clusters. Colouring of the heatmap indicates that clustering of these soil classes is largely based upon high metabolite concentrations in Cluster B^4; medium concentrations in Cluster B^6; and relatively low concentrations in Cluster B^5 (Fig. 3). The lack of significant difference observed between general soil properties (Fig. S8) measured in each of these three clusters indicates that no measured characteristic is solely responsible for this metabolomic variation. Metabolomic differences may therefore be explained by the combined influence of multiple factors, or due to unmeasured characteristics.

334 The significant loading of metabolite 250754 towards Saline 1 and Saline 2 (Fig. S5)
335 indicates that this metabolite is specific to these soils, implying potential relevant function.
336 Identification of unassigned metabolites may shed light on specific microbial functional
337 pathways, or help identify biomarkers indicative of specific environmental conditions.
338 Combining the powers of MS, for empirical formula, with 2D NMR, for structural distinction
339 between isomers, as described by Bingol and Brüschweiler (2017), could also provide a more
340 powerful means of identifying relevant function. Although unidentified metabolites
341 contributed towards class variance, class distribution did not change significantly when all
342 detected metabolites were included in the PCA model (Fig. S4). Metabolomic class separation
343 through this model does not therefore appear limited by our inability to positively identify all
344 the metabolites in a sample.
345 The majority of assigned metabolites (Fig. S9) included phenolics, organic acids, amino
346 acids and sterols. Uncharged organic molecules such as sterols and lipids typically volatilise
347 readily: a requirement for separation by GC (Lin et al., 2006). Non-volatile molecules
348 containing acidic or basic groups can be volatilised through derivatization (Fiehn, 2016) as
349 conducted in our study (Fiehn et al., 2008). Alongside other derivatization agents, *N*-methyl-
350 *N*-(trimethylsilyl)trifluoroacetamide was used due to its compatibility with small molecules
351 exhibiting a broad range of functional groups (Aretz and Meierhofer, 2016; Fiehn, 2016). The
352 lack of identified metabolites containing thiol or extremely basic, charged amine groups,
353 however, indicates that such metabolites may not have been effectively derivatized. This
354 potential omission of these metabolites may have reduced the resolution of the study:
355 undetected metabolites may be influential in soil class separation. Although GC-MS holds
356 advantages over LC-MS and NMR in terms of the size of spectral databases and spectral
357 resolution (Fiehn, 2016; Pan and Raftery, 2007; Sumner et al., 2015; Tang, 2011), parallel

358 metabolomics studies using these techniques would be advised to improve metabolite
359 coverage, as demonstrated by Psychogios et al. (2011).
360
361 *4.3. Sampling considerations*
362 The number of replicates used in this study (n = 5) was below the lower end of
363 recommendations for PCA (Barrett and Kline, 1981; Comrey and Lee, 2016), however, it did
364 reflect typical soil monitoring programmes. Combined with a low sample-to-variable ratio
365 (Osborne and Costello, 2004), the inter-class differences and similarities inferred through PCA
366 may exist due to error induced by model instability. Accumulation of a larger database of
367 samples analysed using the same methods may therefore be beneficial in reducing PCA model
368 error. This may also reduce error introduced due to the inherent variability within each soil
369 type; significant local environmental factors may have impacted metabolomic profile to a
370 greater extent than soil type. To remedy this, it would be preferential to obtain multiple
371 replicates at each sampling site, and sample from more sites per soil type, consistent with
372 current methods in soil molecular ecology (e.g. Docherty et al., 2015; Lauber et al., 2009;
373 Pershina et al., 2018). If the same sites were sampled over time and analysed independently, a
374 reference control would also have been beneficial for quality assurance purposes (Beger et al.,
375 2019). With a view to defining soil quality through metabolomics, collecting samples of each
376 soil type at a range of depths, on different days and in different seasons may reduce further
377 error introduced through temporal variation in root and microbial activity (Fierer et al., 2003;
378 Preston and Basiliko, 2016; Žifčáková et al., 2016). This is also supported by evidence that
379 rhizodeposition, one of the largest inputs of soil C, is highly responsive to the prevailing
380 conditions (Jones et al., 2009).
381
382 *4.4. Can we infer function from metabolomic profiles?*

A major limitation of this type of study is its snapshot approach to analysing metabolomic profiles. It cannot be concluded whether observed metabolite accumulations existed due to enhanced activity of the pathway through which a metabolite was synthesised (due to slowing of the metabolic process occurring immediately post-synthesis) or due to alteration of transport systems into or out of the cell. Further, the accumulation of a metabolite at one step in a metabolic pathway may have been masked by its presence at normal concentrations in any number of other pathways. Complementary analysis of metabolic flux may therefore have generated a better idea of metabolic network dynamics (Aretz and Meierhofer, 2016; Jeong et al., 2017) through methods such as real-time NMR or MS combined with stable isotopes (Ebrahimi et al., 2016; Link et al., 2015). Further, a metabolomic profile alone cannot provide a complete understanding of interacting molecular pathways and their modes of regulation: increase or decrease in metabolite levels cannot definitively infer functional change. Complementary genomic, proteomic or transcriptomic studies (Trauger et al., 2008) may therefore contribute towards a more holistic understanding of soil microbial regulation and function. This could also be supported by metabolomic profiles of the primary inputs of C to the system.

4.5. Could metabolite expression patterns be used to identify microbial stress responses?

Metabolite accumulation in cells may be expected to occur in response to stress, or due to an imbalance between the kinetics of steps throughout a metabolic pathway (Sheldon et al., 2016; Cao et al., 2019). One of the best examples of the former is the production of osmo- and cryo-protectants in response to extreme temperature and moisture conditions (Warren, 2014; Min et al., 2018). In contrast, there are few metabolomic studies on kinetic imbalance in soil, although stoichiometric imbalances in N, P and C supply have been shown to greatly alter the metabolite profile in freshwater sediments (e.g. accumulation of C storage compounds or

organic acids; Brailsford et al., 2019). The lack of significantly high or low values obtained for stress-linked properties in the TO Brown soil (Table 1) may therefore indicate a lack of environmental stress, allowing for optimal microbial metabolism. This is consistent with the high level of available nutrients (due to fertilizer addition), moderate pH, high rates of primary productivity and organic matter turnover at this site (Table 1), indicating that metabolic rate is not limited through respective deficiencies or acidity. As Pearson correlation separates clusters based on metabolic expression patterns, the presence of TO Brown within Cluster A^3 (Fig. 3A) indicates that equivalent metabolic processes may be occurring in all other Cluster A^3 soils. This implies that Gleyic Sand, SO Gley and Podzolic 2 are not exhibiting specific stress responses. Metabolite accumulation in these soils may instead be due to slowing of the more vulnerable steps in metabolic pathways, or due to NO_3^- or P deficiency-induced rate limitation.

Where Group 2 soils are compared with Cluster B^4 (Fig. 3B), significant differences can be observed in pH and DOC (Fig. S8). As DOC also differs significantly between SO Gley and Podzolic 2 (both members of Cluster B^6), but pH does not, pH appears most influential. This may be expected; pH is considered the dominant influencer of soil microbial community assemblage and C use efficiency (Fierer, 2017; Griffiths et al., 2011; Jones et al., 2019). Cluster B^4 and Group 2 also differ in land-use and vegetation cover, consisting of unimproved and improved pasture, respectively, and dominated by *Vaccinium myrtillus*/*Ulex europaeus* and grassland, respectively (Fig. S1). The occurrence of members of Cluster B^4 and Group 2 in separate clusters (Clusters A^2 and A^3, respectively), through Pearson's correlation (Fig. 3A), indicates that a significantly different pattern of metabolite expression is observed in these soils. As Cluster A^3 soils do not appear to be exhibiting specific stress responses (as previously discussed) a pH-, agricultural improvement- or vegetation cover-induced stress response may be occurring in Cluster A^2 soils. The main difference between expression profile shape of Cluster A^2 and Cluster A^3 is the higher concentrations of metabolites Nah to mannose in the

433 former (Fig. 3A). This may suggest a relationship between these metabolites and the implied
434 stress response. An equivalent response is also apparent in NC Lithosols. This may be expected
435 due to equivalent pH (Table 1), lack of agricultural improvement and non-grassland cover,
436 when compared with Cluster A^2 soils.

437 The above discussion highlights that much more work is needed to explore how the
438 metabolome responds to a range of management factors and external stressors. This
439 information can then be used to benchmark soil metabolomic responses. It may also allow us
440 to identify specific biomarkers rather than relying on a fingerprinting approach.

441

442 **5. Conclusions and future perspectives**

443 Based on this study, the inter-class variance between the metabolomic profiles of
444 different soil classes, as defined by GC-MS, is not sufficient to uniquely define soil quality.
445 An increased number of samples per class may improve PCA model stability, however, more
446 accurately distributing variance. By combining this with metabolic flux analysis and
447 complementary metabolomics through LC-MS or NMR, a more robust dataset may be
448 produced, maintaining the potential for metabolomics to gauge soil quality.

449 The number of detected but unassigned metabolites observed in our study emphasises
450 current limitations in terms of metabolite library sizes. Although these did not greatly impact
451 the patterns of inter-class variance, the nature of unassigned metabolites may be significant in
452 broadening current understandings of soil microbial function, or for biomarker discovery.
453 Structure elucidation through targeted MS coupled with NMR may therefore be considered
454 critical for more rigorous metabolomic characterisation.

455 Our results also show that the metabolome may respond to environmental influences
456 such as pH, land-use, moisture and Na content. Coupling the metabolomic profiles of discrete
457 soil classes with measured characteristics has therefore allowed for direction of future studies

through attribution of metabolite expression profiles to soil characteristics and molecular pathways. Enzyme kinetics and binding studies may also allow for identification of specific regulatory mechanisms that dictate metabolite expression associated with function. Combined with genomics, proteomics and transcriptomics, distinction could more readily be made between metabolite-induced enzyme inhibition and genomic or proteomic regulation.

Coupling metabolomics with the described combination of techniques therefore holds great potential to provide an in-depth and holistic understanding of soil microbial molecular pathways and their association with environmental cues. Gaining understanding here may have implications regarding biomolecular dynamics and nutrient cycling linked to ecosystem service provision. The understanding gained through metabolomics and complementary experimental methodologies may therefore provide a basis for management guidelines and direct more sustainable intensification in a functional landscape.

Acknowledgements

This work was carried out under the DOMAINE Project funded by the UK Natural Environment Research Council (NE/K010689/1). Metabolomics analysis was provided by the West Coast Metabolomics Centre.

References

Abram, F., 2015. Systems-based approaches to unravel multi-species microbial community functioning. Computational and Structural Biotechnology Journal 13, 24-32.

Aretz, I., Meierhofer, D., 2016. Advantages and pitfalls of mass spectrometry based metabolome profiling in systems biology. International Journal of Molecular Sciences 17, 632.

Avery, B.W., 1990. Soils of the British Isles. CAB International, Wallingford, UK.

Barberan, A., Bates, S.T., Casamayor, E.O., Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME Journal 6, 343-351.

Barrett, P., Kline, P., 1981. The observation to variable ratio in factor analysis. Personality Study & Group Behaviour 1, 23-33.

Barupal, D.K., Zhang, Y., Shen, T., Fan, S.L., Roberts, B.S., Fitzgerald, P., Wancewicz, B., Valdiviez, L., Wohlgemuth, G., Byram, G., Choy, Y.Y., Haffner, B., Showalter, M.R., Vaniya, A., Bloszies, C.S., Folz, J.S., Kind, T., Flenniken, A.M., McKerlie, C., Nutter, L.M.J., Lloyd, K.C., Fiehn, O., 2019. A comprehensive plasma metabolomics dataset for a cohort of mouse knockouts within the international mouse phenotyping consortium. Metabolites 9, 101.

Beger R.D., Dunn, W.B., Bandukwala, A., Bethan, B., Broadhurst, D., Clish, C.B., Dasari, S., Derr, L., Evans, A., Fischer, S., Flynn, T., Hartung, T., Herrington, D., Higashi, R., Hsu, P.C., Jones, C., Kachman, M., Karuso, H., Kruppa, G., Lippa, K., Maruvada, P., Mosley, J., Ntai, I., O'Donovan, C., Playdon, M., Raftery, D., Shaughnessy, D., Souza, A., Spaeder, T., Spalholz, B., Tayyari, F., Ubhi, B., Verma, M., Walk, T., Wilson, I., Witkin, K., Bearden, D.W., Zanetti, K.A., 2019. Towards quality assurance and quality control in untargeted metabolomics studies. Metabolomics 15, 4.

Bellamy, P.H., Loveland, P.J., Bradley, R.I., Lark, R.M., Kirk, G.J.D., 2005. Carbon losses from all soils across England and Wales 1978-2003. Nature 437, 245-248.

Bending, G.D., Turner, M.K., Rayns, F., Marx, M.-C., Wood, M., 2004. Microbial and biochemical soil quality indicators and their potential for differentiating areas under contrasting agricultural management regimes. Soil Biology and Biochemistry 36, 1785-1792.

506 Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A.,
507 Knight, R., Fierer, N., 2011. The under-recognized dominance of Verrucomicrobia in
508 soil bacterial communities. Soil Biology & Biochemistry 43, 1450-1455.
509 Bingol, K., Bruschweiler-Li, L., Li, D., Zhang, B., Xie, M., Brüschweiler, R., 2016. Emerging
510 new strategies for successful metabolite identification in metabolomics. Bioanalysis 8,
511 557-573.
512 Bingol, K., Brüschweiler, R., 2017. Knowns and unknowns in metabolomics identified by
513 multidimensional NMR and hybrid MS/NMR methods. Current Opinion in
514 Biotechnology 43, 17-24.
515 Brailsford, F.L., Glanville, H.C., Golyshin, P.N., Marshall, M.R., Lloyd, C.E., Johnes, P.J.,
516 Jones, D.L., 2019. Nutrient enrichment induces a shift in dissolved organic carbon (DOC)
517 metabolism in oligotrophic freshwater sediments. Science of the Total Environment 690,
518 1131-1139.
519 Bundy, J.G., Davey, M.P., Viant, M.R., 2009. Environmental metabolomics: a critical review
520 and future perspectives. Metabolomics 5, 3-21.
521 Bundy, J.G., Ramløv, H., Holmstrup, M., 2003. Multivariate metabolic profiling using ^{1}H
522 nuclear magnetic resonance spectroscopy of freeze-tolerant and freeze-intolerant
523 earthworms exposed to frost. Cryo-Letters 24, 347-358.
524 Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens,
525 L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W.,
526 Brussaard, L., 2018. Soil quality - A critical review. Soil Biology and Biochemistry 120,
527 105-125.
528 Cao, Y.W., Qu, R.J., Miao, Y.J., Tang, X.Q., Zhou, Y., Wang, L., Geng, L., 2019. Untargeted
529 liquid chromatography coupled with mass spectrometry reveals metabolic changes in
530 nitrogen-deficient *Isatis indigotica* Fortune. Phytochemistry 166, 112058.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S., Xia, J., 2018. MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. Nucleic Acids Research 46, W486-W494.

Comrey, A.L., Lee, H.B., 2016. A First Course in Factor Analysis. Psychology Press Ltd., Hove, UK.

Docherty, K.M., Borton, H.M., Espinosa, N., Gebhardt, M., Gil-Loaiza, J., Gutknecht, J.L.M., Maes, P.W., Mott, B.M., Parnell, J.J., Purdy, G., Rodrigues, P.A.P., Stanish, L.F., Walser, O.N., Gallery, R.E., 2015. Key edaphic properties largely explain temporal and geographic variation in soil microbial communities across four biomes. PLOS ONE 10, e0135352.

Ebrahimi, P., Larsen, F.H., Jensen, H.M., Vogensen, F.K., Engelsen, S.B., 2016. Real-time metabolomic analysis of lactic acid bacteria as monitored by in vitro NMR and chemometrics. Metabolomics 12, UNSP 77.

Emmett, B.A., Frogbrook, Z.L., Chamberlain, P.M., Griffiths, R., Pickup, R., Poskitt, J., Reynolds, B., Rowe, E., Rowland, P., Spurgeon, D., Wilson, J., Wood, C.M., 2008. Countryside Survey. Soils Manual. NERC/Centre for Ecology & Hydrology, CS Technical Report No.3/07, CEH Project Number: C03259, 180 pp.

Farrell, M., Macdonald, L.M., Hill, P.W., Wanniarachchi, S.D., Farrar, J., Bardgett, R.D., Jones, D.L., 2014. Amino acid dynamics across a grassland altitudinal gradient. Soil Biology & Biochemistry 76, 179-182.

Fiehn, O., 2016. Metabolomics by gas chromatography-mass spectrometry: combined targeted and untargeted profiling. Current Protocols in Molecular Biology 114, 30.4.1-30.4.32.

Fiehn, O., Kopka, J., Trethewey, R.N., Willmitzer, L., 2002. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Analytical Chemistry 72, 3573-3580.

556 Fiehn, O., Wohlgemuth, G., Scholz, M., Kind, T., Lee, D.Y., Lu, Y., Moon, S., Nikolau, B.,
557 2008. Quality control for plant metabolomics: Reporting MSI-compliant studies. Plant
558 Journal 53, 691-704.
559 Fierer, N., 2017. Embracing the unknown: disentangling the complexities of the soil
560 microbiome. Nature Reviews Microbiology 15, 579-590.
561 Fierer, N., Schimel, J.P., Holden, P.A., 2003. Variations in microbial community composition
562 through two soil depth profiles. Soil Biology & Biochemistry 35, 167-176.
563 Fu, J., Gong, Z., Kelly, B.C., 2019. Metabolomic profiling of zebrafish (*Danio rerio*) embryos
564 exposed to the antibacterial agent triclosan. Environmental Toxicology and Chemistry 38,
565 240-249.
566 George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., Griffiths, R.I.,
567 Lebron, I., Emmett, B.A., Robinson, D.A., Jones, D.L., 2019. Divergent national-scale
568 trends of microbial and animal biodiversity revealed across diverse temperate soil
569 ecosystems. Nature Communications 10, 1107.
570 Graham, E.B., Crump, A.R., Kennedy, D.W., Arntzen, E., Fansler, S., Purvine, S.O., Nicora,
571 C.D., Nelson, W., Tfaily, M.M., Stegen, J.C., 2018. Multi 'omics comparison reveals
572 metabolome biochemistry, not microbiome composition or gene expression, corresponds
573 to elevated biogeochemical function in the hyporheic zone. Science of The Total
574 Environment 642, 742-753.
575 Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., Whiteley, A.S., 2011. The
576 bacterial biogeography of British soils. Environmental Microbiology 13, 1642-1654.
577 Guijas, C., Montenegro-Burke, J.R., Warth, B., Spilker, M.E., Siuzdak, G., 2018.
578 Metabolomics activity screening for identifying metabolites that modulate phenotype.
579 Nature Biotechnology 36, 316-320.

580 Huan, T., Troyer, D.A., Li., L., 2016. Metabolite analysis and histology on the exact same
581 tissue: comprehensive metabolomic profiling and metabolic classification of prostate
582 cancer. Scientific Reports 6, 32272.
583 Jeong, S., Eskandari, R., Park, S.M., Alvarez, J., Tee, S.S., Weissleder, R., Kharas, M.G., Lee,
584 H., Keshari, K.R., 2017. Real-time quantitative analysis of metabolic flux in live cells
585 using a hyperpolarized micromagnetic resonance spectrometer. Science Advances 3,
586 e1700341.
587 Jones, D.L., Owen, A.G., Farrar, J.F., 2002. Simple method to enable the high resolution
588 determination of total free amino acids in soil solutions and soil extracts. Soil Biology &
589 Biochemistry 34, 1893-1902.
590 Jones, D.L., Nguyen, C., Finlay, R.D., 2019. Carbon flow in the rhizosphere: carbon trading at
591 the soil-root interface. Plant and Soil 321, 5-33.
592 Jones, D.L., Hill, P.W., Smith, A.R., Farrell, M., Ge, T., Banning, N.C., Murphy, D.V., 2018.
593 Role of substrate supply on microbial carbon use efficiency and its role in interpreting
594 soil microbial community-level physiological profiles (CLPP). Soil Biology &
595 Biochemistry 123, 1-6.
596 Jones, D.L., Cooledge, E.C., Hoyle, F.C., Griffiths, R.I., Murphy, D.V., 2019. pH and
597 exchangeable aluminum are major regulators of microbial energy flow and carbon use
598 efficiency in soil microbial communities. Soil Biology & Biochemistry *in press.*
599 Jones, O.A.H., Maguire, M.L., Griffin, J.L., Dias, D.A., Spurgeon, D.J., Svendsen, C., 2013.
600 Metabolomics and its use in ecology. Austral Ecology 38, 713-720.
601 Jones, O.A.H., Sdepanian, S., Lofts, S., Svendsen, C., Spurgeon, D.J., Maguire, M.L., Griffin,
602 J.L., 2014. Metabolomic analysis of soil communities can be used for pollution
603 assessment. Environmental Toxicology and Chemistry 33, 61-64.

Jurburg, S.D., Salles, J.F., 2015. Functional redundancy and ecosystem function - the soil microbiota as a case study. In: Biodiversity in Ecosystems - Linking Structure and Function (Eds. Blanco, J.., Lo, Y.H., Roy, S.), IntechOpen Ltd., London, UK,

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental Scale. Applied and Environmental Microbiology 75, 5111-5120.

Lin, C.Y., Viant, M.R., Tjeerdema, R.S., 2006. Metabolomics: Methodologies and applications in the environmental sciences. Journal of Pesticide Science 31, 245-251

Link, H., Fuhrer, T., Gerosa, L., Zamboni, N., Sauer, U., 2015. Real-time metabolome profiling of the metabolic switch between starvation and growth. Nature Methods 12, 1091-1097.

Mahieu, N.G., Patti, G.J., 2017. Systems-level annotation of a metabolomics data set reduces 25000 features to fewer than 1000 unique metabolites. Analytical Chemistry 89, 10397-10406.

Min, K., Showman, L., Perera, A., Arora, R., 2018. Salicylic acid-induced freezing tolerance in spinach (*Spinacia oleracea* L.) leaves explored through metabolite profiling. Environmental and Experimental Botany 156, 214-227.

Miranda, K.M., Espey, M.G., Wink, D.A., 2001. A rapid, simple spectrophotometric method for simultaneous detection of nitrate and nitrite. Nitric Oxide 5, 62-71.

Mulvaney, R.L., 1996. Nitrogen - inorganic forms. In: Methods of Soil Analysis, Part 3, (Ed. Sparks, D.L.). Soil Science Society of America, American Society of Agronomy, Madison, WI, USA.

Murphy, J., Riley, J.P., 1962. A modified single solution method for determination of phosphate in natural waters. Analytica Chimica Acta 27, 31-36.

Osborne, J.W., Costello, A.B., 2004. Sample size and subject to item ratio in principal components analysis. Practical Assessment, Research and Evaluation 9, 8.

629 Pan, Z., Raftery, D., 2007. Comparing and combining NMR spectroscopy and mass
630 spectrometry in metabolomics. Analytical and Bioanalytical Chemistry 387, 525-527.
631 Patti, G.J., Yanes, O., Siuzdak, G., 2012. Metabolomics: the apogee of the omics trilogy.
632 Nature Reviews Molecular Cell Biology 13, 263-269.
633 Pershina, E. V., Ivanova, E.A., Korvigo, I.O., Chirak, E.L., Sergaliev, N.H., Abakumov, E. V.,
634 Provorov, N.A., Andronov, E.E., 2018. Investigation of the core microbiome in main soil
635 types from the East European plain. Science of the Total Environment 631-632, 1421-
636 1430.
637 Preston, M.D., Basiliko, N., 2016. Carbon mineralization in peatlands: Does the soil microbial
638 community composition matter? Geomicrobiology Journal 33, 151-162.
639 Psychogios, N., Hau, D.D., Peng, J., Guo, A.C., Mandal, R., Bouatra, S., Sinelnikov, I.,
640 Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang,
641 P., Hollander, Z., Pedersen, T.L., Smith, S.R., Bamforth, F., Greiner, R., McManus, B.,
642 Newman, J.W., Goodfriend, T., Wishart, D.S., 2011. The human serum metabolome.
643 PLoS ONE 6, e16957.
644 Schloter, M., Nannipieri, P., Sørensen, S.J., van Elsas, J.D., 2018. Microbial indicators for soil
645 quality. Biology and Fertility of Soils 54, 1-10.
646 Seaton, F.M., Barrett G., Burden, A., Creer, S., Dos Santos Pereira, G., Fitos, E., Garbutt, A.,
647 Griffiths, R.I., Henrys, P., Jones, D.L., Keith, A., Lebron, I., Maskell, L., Reinsch, S.,
648 Smart, S.M., Williams, B., Emmett, B., Robinson, D.A., 2019. Identifying soil functional
649 classes from monitoring data at a national scale. European Journal of Soil Science
650 *submitted.*
651 Shelden, M.C., Dias, D.A., Jayasinghe, N.S., Bacic, A., Roessner, U., 2016. Root spatial
652 metabolite profiling of two genotypes of barley (*Hordeum vulgare* L.) reveals differences
653 in response to short-term salt stress. Journal of Experimental Botany 67, 3731-3745.

654 Simfukwe, P., Hill, P.W., Emmett, B.A., Jones, D.L., 2011. Soil classification provides a poor
655 indicator of carbon turnover rates in soil. Soil Biology & Biochemistry 43, 1688-1696.
656 Sumner, L.W., Lei, Z., Nikolau, B.J., Saito, K., 2015. Modern plant metabolomics: advanced
657 natural product gene discoveries, improved technologies, and future prospects. Natural
658 Products Reports 32, 212-219.
659 Swenson, T.L., Jenkins, S., Bowen, B.P., Northen, T.R., 2015. Untargeted soil metabolomics
660 methods for analysis of extractable organic matter. Soil Biology & Biochemistry 80, 189-
661 198.
662 Tang, J., 2011. Microbial metabolomics. Current Genomics 12, 391-403.
663 Trauger, S.A., Kalisak, E., Kalisiak, J., Morita, H., Weinberg, M. V., Menon, A.L., Poole II,
664 F.L., Adams, M.W.W., Siuzdak, G., 2008. Correlating the transcriptome, proteome, and
665 metabolome in the environmental adaptation of a hyperthermophile. Journal of Proteome
666 Research 7, 1027-1035.
667 Vinayavekhin, N., Saghatelian, A., 2010. Untargeted metabolomics. Current Protocols in
668 Molecular Biology 90, 30.1.1-30.1.24.
669 Wang, X., Yang, F., Zhang, Y., Xu, G., Liu, Y., Tian, J., Gao, P., 2015. Evaluation and
670 optimization of sample preparation methods for metabolic profiling analysis of
671 *Escherichia coli*. Electrophoresis 36, 2140-2147.
672 Warren, C.R., 2014. Response of osmolytes in soil to drying and rewetting. Soil Biology &
673 Biochemistry 70, 22-32.
674 Wellerdiek, M., Winterhoff, D., Reule, W., Brandner, J., Oldiges, M., 2009. Metabolic
675 quenching of *Corynebacterium glutamicum*: efficiency of methods and impact of cold
676 shock. Bioprocess and Biosystems Engineering 32, 581-592.
677 Wilson, I.D., Plumb, R., Granger, J., Major, H., Williams, R., Lenz, E.M., 2005. HPLC-MS-
678 based methods for the study of metabolomics. Journal of Chromatography B 817, 67-76.

679 Wishart, D.S., 2019. Metabolomics for investigating physiological and pathophysiological
680 processes. Physiological Reviews 99, 1819-1875.
681 Zhao, L.J., Zhang, H.L., White, J.C., Chen, X.Q., Li, H.B., Qu, X.L., Ji, R., 2019.
682 Metabolomics reveals that engineered nanomaterial exposure in soil alters both soil
683 rhizosphere metabolite profiles and maize metabolic pathways. Environmental Science
684 Nano 6, 1716-1727.
685 Žifčáková, L., Větrovský, T., Howe, A., Baldrian, P., 2016. Microbial activity in forest soil
686 reflects the changes in ecosystem properties between summer and winter. Environmental
687 Microbiology 18, 288-301.